On Random Network Coding for Multicast

Adrian Tauste Campo
Universitat Pompeu Fabra
Barcelona, Spain

Alex Grant
Institute for Telecommunications Research
University of South Australia

***Abstract*— Random linear network coding is a particularly decentralized approach to the multicast problem. Use of random network codes introduces a non-zero probability however that some sinks will not be able to successfully decode the required sources. One of the main theoretical motivations for random network codes stems from the lower bound on the probability of successful decoding reported by Ho et. al. (2003). This result demonstrates that all sinks in a linearly solvable network can successfully decode all sources provided that the random code field size is large enough. This paper develops a new bound on the probability of successful decoding.**

I. Introduction

It has been recently proved that network layer coding can increase throughput, particularly for multicast scenarios [1]. It is also known that linear network codes [2] can achieve max-flow upper bounds on the throughput in a single source multicast network. The algebraic approach of [3] is particularly useful in the design and analysis of linear network codes, and we adopt the notation and terminology of that paper.

Random networks codes [4], [5] are linear network codes in which the encoding coefficients are chosen randomly from a finite field. The sink nodes can decode correctly if and only if the overall transfer matrix from the sources to each sink is invertible. One of the main theoretical results for random network codes consists of the following lower bound on the probability of successful decoding [4], assuming that the underlying network is linearly solvable over $\mathbb{F}_q$ (i.e. there exists a linear code which satisfies the multicast requirements). For a network code in which some of the code coefficients are chosen independently and uniformly from a finite field with cardinality q, the probability that all d receivers can decode the source processes is at least

$$\left(1-\frac{d}{q}\right)^{\nu} \tag{1}$$

where ν is the maximum number of links receiving signals with independent random coefficients in any set of links constituting a flow solution from all sources to any receiver [5].

A looser bound (subject to the same conditions as above) which depends only on η, the total number of edges receiving signals with independent random coefficients is given by [4], [6]

$$\left(1-\frac{d}{q}\right)^{\eta}. \tag{2}$$

Thus provided a linear solution over $\mathbb{F}_q$ exists in the first place, the probability of successful decoding can be made as close to one as desired, by increasing the field size q. The bounds (1) and (2) rely on the special structure of the determinant polynomial of the transfer matrix of the network.

This paper develops the following new lower bound.

Theorem 1: Consider a network code in which η edges receive signals with independent random coefficients chosen independently and uniformly from a finite field with cardinality q. If there is some choice of coefficients for these η edges that results in a solution over $\mathbb{F}_q$ then the probability that all receivers can decode the source processes is at least

$$\left(1-\frac{1}{q}\right)^{\eta}. \tag{3}$$

Our approach for the proof of this theorem is to identify a critical sub-matrix of the Edmonds matrix whose non-singularity is a necessary and sufficient condition for decoding success. This critical matrix is different for each sink in the network. The new bound results directly from a nesting property of the critical matrices.

In the new bound, the field size q required to attain a given probability of success depends only on the number of edges with random coefficients, and not on the number of sinks. The resulting d-fold reduction in the required q could be significant. We emphasize that (3), like (1) applies only when the underlying network is solvable over $\mathbb{F}_q$. This is a consequence of the conditions for applicability of the Schwartz-Zippel inequality, which is used in the proof of both bounds. Thus (3) does *not* imply the universal existence of binary solutions for every network. The bounds (1), (2) and (3) only provide lower bounds for a given q when the network is solvable over $\mathbb{F}_q$.

We further conjecture that for large random networks satisfying certain properties, the success probability behaves as

$$\prod_{i=1}^{E}\left(1-\frac{1}{q^i}\right) \tag{4}$$

where E is the total number of links in the network.

The paper is organized as follows: Section II presents our model and introduces some algebraic notation. Section III develops the new bound (3), while Section V discusses random graphs, leading to the conjecture (4).

II. Network Coding Model

We adopt the model from [3]. The network is represented by a directed acyclic graph $\mathcal{G}=(\mathcal{V},\mathcal{E})$ with $V=|\mathcal{V}|$ nodes and $E=|\mathcal{E}|$ edges. There are r independent, discrete source

processes with messages belonging to $\mathbb{F}_q$, and $d \geq 1$ receivers. Each receiver node has $L \geq r$ incoming edges. The multicast requirement is that each receiver node can decode every source message from the signals on its incident edges.

Each edge $e \in \mathcal{E}$ is incident to node $v \in \mathcal{V}$ if $v = \mathrm{head}(e)$, or is an outgoing edge if $v = \mathrm{tail}(l)$. The in-degree of a node v is $d_{\text{in}}(v)$ and the out-degree is $d_{\text{out}}(v)$. The time unit is chosen such that the capacity of each link is one bit per unit time and edges with larger capacity are modeled as parallel edges. Without loss of generality, it can be assumed that each source is associated with a source node $s_\alpha \in \mathcal{V}$ with $d_{\text{in}}(s_\alpha) = 0$ and $d_{\text{out}}(s_\alpha) = 1$, $\alpha = 1, 2, \ldots, r$. Similarly, each sink node t_β has $d_{\text{in}}(t_\beta) = r$ and $d_{\text{out}}(t_\beta) = 0$, $\beta = 1, 2, \ldots, d$ (it is always possible to obtain such a graph by introducing auxiliary nodes and edges). It will further be assumed that edges are labeled ancestrally.

A *scalar linear network code* for G is an assignment of linear encoding functions $f_v : \mathbb{F}_q^{d_{\text{in}}(v)} \mapsto \mathbb{F}_q^{d_{\text{out}}(v)}$ to each node $v \in \mathcal{V}$. Such codes are sufficient for the multicast problem on acyclic delay networks. Following [3], define the *encoding matrix* $F \in \mathbb{F}_q^{E \times E}$ where F_{ij} is the coefficient applied to the symbol incoming on edge $i \in \mathcal{E}$ for contribution to outgoing edge $j \in \mathcal{E}$. According to the assumption of ancestral ordering, F is strictly upper triangular. Similarly, the *source matrix* $A \in \mathbb{F}_q^{r \times E}$ maps messages onto outgoing source edges and the *sink matrix* $B_\beta \in \mathbb{F}_q^{r \times E}$ maps incoming sink edges onto the sinks $t_\beta \in \mathcal{V}$, $\beta = 1, 2, \ldots, d$.

Let $x \in \mathbb{F}_q^{1 \times r}$ be a row vector representing the source messages. Then the received vector of symbols $y_\beta \in \mathbb{F}_q^{1 \times r}$ at sink $\beta = 1, 2, \ldots, d$ is given by

$$y_\beta = x M_\beta$$

where

$$M_\beta = A(I - F)^{-1} B_\beta^T.$$

Each sink can decode all sources if and only if $\det(A(I - F)^{-1} B_\beta^T) \neq 0$ for every $\beta = 1, 2, \ldots, d$, or equivalently if the Edmonds matrix

$$Z_\beta = \begin{bmatrix} A & 0 \\ I - F & B_\beta^T \end{bmatrix}$$

is non-singular.

Considering the entries of A, F and B_β as variables, the Leibniz determinant formula provides a way of writing $\det Z_\beta$ as a multivariate polynomial P_β in the a_{ij}, f_{ij}, b_{ij}. Furthermore, this multivariate polynomial has degree at most ν but is linear in each variable individually. Therefore the product

$$P = \prod_\beta P_\beta \tag{5}$$

has degree $d\nu$, with each variable of degree d or less.

The lower bound (1) results from a modified Schwartz-Zippel bound, which takes into account the individual variable degree constraint of P_β [5, Lemma 1]. We reproduce this lemma for reference.

Lemma 1: Let P be a multivariate polynomial of degree $d\nu$, with the exponent of any individual variable at most d. Let each variable be chosen uniformly from $\mathbb{F}_q$. Then if P is not identically zero,

$$\Pr(P \neq 0) \geq \left(1 - \frac{d}{q}\right)^\nu. \tag{6}$$

We make two remarks on this approach. First, application of Lemma 1 to P as defined in (5) implies an independence of the events $P_{\beta_1} = 0$ and $P_{\beta_2} = 0$. Depending on the structure of the network, these events may be strongly dependent. For example, consider $P_1 = P_2 = \cdots = P_d$, meaning all sinks have identical incoming signals ($B_1 = B_2 = \cdots = B_d$). Then Lemma 1 yields a lower bound $(1 - d/q)^\nu$, rather than $(1 - 1/q)^\nu$. Obviously this is an extreme example, yet it illustrates the point that (1) may be loose.

Secondly, the modified Schwartz-Zippel bound itself can be very loose, as the following example shows. Let $H \in \mathbb{F}_q^{m \times m}$ with each entry h_{ij} chosen independently with a uniform distribution on $\mathbb{F}_q$. Then it is well known that

$$\Pr(\det H \neq 0) = \pi_m(q) = \prod_{i=1}^{m} \left(1 - q^{-i}\right). \tag{7}$$

In contrast, Lemma 1 gives the lower bound

$$\Pr(\det H \neq 0) \geq \left(1 - q^{-1}\right)^m, \tag{8}$$

which also could be obtained from (7) by lower bounding each term in the product by the minimum term $(1 - q^{-1})$.

We emphasize that (6) applies only when P is not identically zero for every choice of variables (e.g. all coefficients are zero). This precludes application of (6) to non-solvable networks, i.e. networks where every choice of F makes Z_β singular and hence $P = 0$.

In Section III we partially address the dependency between the P_β, while in Section V we consider large random networks, where we also discuss the extent to which (7) improves (8).

III. The New Bound

According to our assumption regarding sources and sinks, and the ancestral ordering of edges, we can further assume without loss of generality that

$$A = \begin{bmatrix} I_{r \times r} & 0_{r \times (E-r)} \end{bmatrix}$$
$$B_\beta = \begin{bmatrix} 0_{r \times k_\beta} & I_{r \times r} & 0_{r \times (E-r-k_\beta)} \end{bmatrix}, \beta = 1, 2, \ldots, d$$

where $k_1 > r$ and $k_\beta > r + k_{\beta-1}$, $\beta > 1$. This means that the sources inject messages into the network via edges $1, 2, \ldots, r$ and that each sink observes signals on r consecutively numbered edges. No sink shares edges with any other sink or source. See Figure 1 for an example of how to arrive at this formulation.

Then the Edmonds matrix for sink β has the following structure:

$$Z_\beta = \begin{bmatrix} I_r & 0 & 0 & 0 & 0 \\ U_1 & W_{11} & W_{12} & W_{13} & 0 \\ 0 & U_2 & W_{21} & W_{22} & 0 \\ 0 & 0 & U_3 & W_{31} & I_r \\ 0 & 0 & 0 & U_4 & 0 \end{bmatrix} \tag{9}$$

where the U_i are square, upper triangular with diagonal elements all equal to 1. The matrices U_1 and U_3 are $r \times r$, U_2 is $(k_\beta - 2r) \times (k_\beta - 2r)$ and U_4 is $(E - r - k_\beta) \times (E - r - k_\beta)$.

Definition 1: The *critical matrix* for sink β is the following $(k_\beta - r) \times (k_\beta - r)$ principal sub-matrix of Z_β,

$$C_\beta = \begin{pmatrix} W_{11} & W_{12} \\ U_2 & W_{21} \end{pmatrix}. \tag{10}$$

Lemma 2: The determinant of the Edmonds matrix for sink β has the same magnitude as the determinant of its critical matrix.

$$|\det Z_\beta| = |\det C_\beta|$$

Proof: Straightforward from either the Laplace expansion of $\det Z_\beta$, or repeated application of the partitioned matrix determinant formula. ∎

We can immediately apply Lemma 1 to $\det C_\beta$ to bound the probability for a given sink

$$\Pr(\det Z_\beta \neq 0) = \Pr(\det C_\beta \neq 0) \geq \left(1 - \frac{1}{q}\right)^{\eta_\beta}, \tag{11}$$

where η_β is the number of columns in C_β with variable terms, i.e. the number of edges in the subset $\{r+1, r+2, \ldots, k_\beta\}$ receiving signals with random coefficients.

For the d receiver problem, we have the following very useful property of the critical matrices, which is guaranteed by their construction.

Lemma 3 (Nesting of critical matrices): C_{β_1} is a principal sub-matrix of C_{β_2} for $\beta_2 > \beta_1$.

Hence each critical matrix C_β has as nested principal sub-matrices, all the critical matrices for sinks $1, 2, \ldots, \beta - 1$.

Proof: [Proof of main result (3)] Let E_β, $\beta = 1, 2, \ldots, d$ be the event that sink β can decode. By Lemma 2, $\mathsf{E}_\beta \iff \det Z_\beta \neq 0 \iff \det C_\beta \neq 0$. Now the probability that all sinks can decode is given by

$$\Pr\left(\bigcap_{\beta=1}^{d} \mathsf{E}_\beta\right) = \Pr(\mathsf{E}_1) \Pr(\mathsf{E}_2 \mid \mathsf{E}_1) \ldots \Pr(\mathsf{E}_\beta \mid \mathsf{E}_1 \ldots \mathsf{E}_{\beta-1}) \tag{12}$$

Now consider $\Pr(\mathsf{E}_m \mid \mathsf{E}_1, \ldots, \mathsf{E}_{m-1}) = \Pr(\det C_m \neq 0 \mid \det C_1 \neq 0, \ldots, \det C_{m-1} \neq 0)$ for some $2 \leq m \leq \beta$. By Lemma 3, C_m can be partitioned

$$C_m = \begin{pmatrix} C_{m-1} & U \\ V & W \end{pmatrix}$$

for appropriate choices of U, V, W.

Conditioned on $\det C_{m-1} \neq 0$, we can use the partitioned matrix determinant formula to write

$$\det C_m = \det(C_{m-1}) \det\left(W - V C_{m-1}^{-1} U\right), \tag{13}$$

which (conditioned on $\det C_{m-1} \neq 0$) is zero if and only if $\det\left(W - V C_{m-1}^{-1} U\right) = 0$.

Let ϕ_m be the multivariate polynomial corresponding to $\det C_m$, and let σ_{m-1} be the multivariate polynomial corresponding to $\det\left(W - V C_{m-1}^{-1} U\right)$. Then from (13) $\deg \phi_m = \deg \phi_{m-1} + \deg \sigma_{m-1}$. This relation also holds for the degree of any individual variable. From the Leibniz formula and the structure of the Edmonds matrix (as explained previously for P_β), we also know that the individual degree of any variable in ϕ_m or ϕ_{m-1} is zero or one. Hence

$$\deg \sigma_{m-1} = \deg \phi_m - \deg \phi_{m-1},$$

and the degree of any individual variable in σ_{m-1} is at most 1. Collecting results so far and applying Lemma 1,

$$\begin{aligned} \Pr(\mathsf{E}_m \mid \mathsf{E}_1, \ldots, \mathsf{E}_{m-1}) &= \Pr\left(\det\left(W - V C_{m-1}^{-1} U\right) \neq 0\right) \\ &= \Pr(\sigma_{m-1} \neq 0) \\ &\leq \left(1 - \frac{1}{q}\right)^{\deg \phi_m - \deg \phi_{m-1}} \end{aligned}$$

Finally, substitution into (12) results in a telescoping sum for the exponents, $\deg \phi_1 + \deg \phi_2 - \deg \phi_1 + \deg \phi_3 - \deg \phi_2 + \ldots$, leaving only

$$\Pr\left(\bigcap_{\beta=1}^{d} \mathsf{E}_\beta\right) \geq \left(1 - \frac{1}{q}\right)^{\deg \phi_d}$$

This directly yields (3) via $d\nu \leq \eta \triangleq \deg \phi_d = \eta_d \leq E$. ∎

Let

$$z(d, q) = \frac{\log(1 - d/q)}{\log(1 - 1/q)}.$$

Then (3) is tighter than (1) whenever

$$\eta < \nu\, z(d, q).$$

Furthermore, $z(d, q) > d$ and

$$\lim_{q \to d} z(d, q) = \infty$$
$$\lim_{q \to \infty} z(d, q) = d.$$

Roughly speaking, the new bound is tighter for networks with $E = O(\nu d)$ and sufficiently small q.

In some instances it may be useful to have a bound which depends only on the total number of edges carrying signals with random coefficients. Replacing ν with η in (1) results in (2) which is looser than (3), since

$$(1 - d/q)^\eta < (1 - 1/q)^\eta.$$

Note that successful decoding at a particular sink β in general depends on only part of C_β. There can be a much smaller sub-matrix that determines singularity, for example, C_β might be block diagonal, with successful decoding of sink β depending only on one of the blocks (this case arises when there are disjoint paths from the sources to each sink). Thus C_β may be larger than strictly required for analysis of sink β alone, however defining the critical matrix this way yields the nesting property that results in the new bound.

IV. Example: The Butterfly Network

Figure 1 shows the well-known butterfly network, with additional nodes and edges introduced in order to satisfy our assumptions on sources and sinks. The source s has $r = 2$ messages, and the edge labels indicate the edge ordering. Edges 1 and 2 carry the two messages from the source, while edges 12 resp. 13 duplicate the signals on edges 5 resp. 10, and edges 14 resp. 15 duplicate 8 resp. 11. Supposing that all other edges carry random linear combinations of signals, $\nu = 7$ and $\eta = 9$.

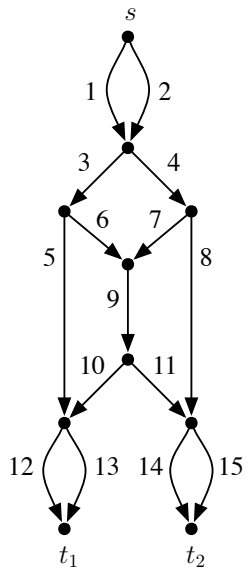

Fig. 1. The butterfly network.

Figure 2 shows the structure of the Edmonds matrix Z_1, and the nested critical matrices C_1 and C_2. To see how the nesting arises, B_2 has been placed alongside. For clarity, most of the zeros have been omitted from each matrix. The solid disks represent random entries of F.

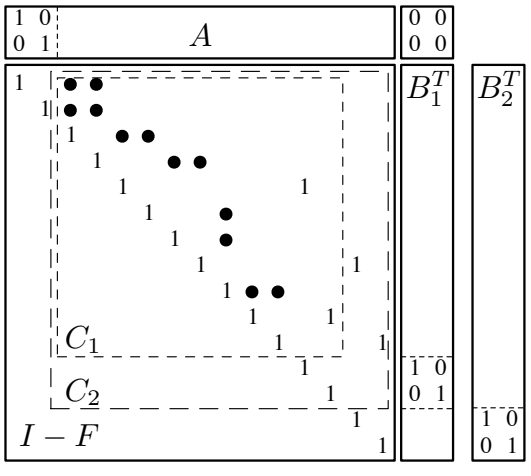

Fig. 2. Critical matrices for the butterfly network.

Figure 3 shows the empirically measured probability of decoding success versus the field size q for the network of Figure 1 (filled circles). This was achieved using monte-carlo simulation, selecting each of the coefficients uniformly from $\mathbb{F}_q$. Results for the first ten prime fields are shown. Also shown are the existing bounds (1), dashed line, (2), solid line, and the new bound (3), dot-dashed line. In this case, the new bound is considerably tighter.

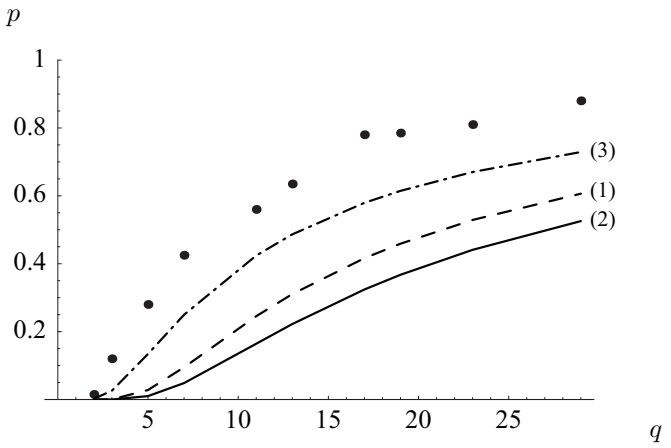

Fig. 3. Success probability p versus field size q compared to bounds (1), (2) and (3) for the butterfly network.

V. Random Graphs

Successful decoding for a particular sink β depends on the non-singularity of its critical matrix C_β. To obtain (3) we used Lemma 1 to bound the probability that this matrix is non-singular. It is interesting to consider however circumstances under which (7) might be applicable, providing an even tighter bound.

There are two main obstacles to the application of (7) for determination of the probability that $\det C_\beta \neq 0$. Firstly, (7) applies to "full" matrices, with each element chosen independently and uniformly from $\mathbb{F}_q$. In contrast, C_β is of the form (10), with all elements below the r-th diagonal equal to zero (the strictly lower triangular part of U_2). Secondly, the non-zero elements in the upper portion (upper triangular part of U_2 and all of W_{11}, W_{12} and W_{21}) of C_β are determined by the topology of the network itself. For a sparsely connected network, the proportion of zeros in this part of the matrix will greatly exceed $1/q$.

Assuming that the random network code coefficients are chosen from the non-zero elements of $\mathbb{F}_q$, the total number of non-zero elements in F is

$$\sigma \triangleq \sum_{v \in \mathcal{V}} d_{\text{in}}(v)\, d_{\text{out}}(v) \leq E^2.$$

Let $\rho = \sigma/E^2$ be the proportion of non-zero elements. Ignoring the structure required by (10), generate a random $m \times m$ matrix $C^{(m)}$ with elements identically distributed according to

$$\Pr(c_{ij} = f) = \begin{cases} 1-\rho & f = 0 \\ \frac{\rho}{q-1} & f \neq 0 \end{cases}$$

It is a remarkable fact that provided ρ does not tend to zero or one too quickly with m,

$$\lim_{m\to\infty} \Pr\left(\det C^{(m)} \neq 0\right) = \pi_m(q).$$

See [7] for a discussion of this threshold effect. Conditioned on the event that $C^{(m)}$ has no all-zero rows or columns (if it did, the network flow would anyway be infeasible regardless of choice of code), the requirement is

$$\rho > \frac{1}{m}\left(\frac{1}{2}\log m + \log\log m\right).$$

This result even holds for independent, but non-identically distributed entries, as discussed by Cooper [7].

Now for sufficiently small ρ, $C^{(m)}$ can be permuted with high probability into the form (10). This leads us to conjecture that there exist conditions on σ such that $\pi_m(q)$ is the success probability for a large, randomly generated network with a given degree distribution. The remainder of this section analyzes some properties of $\pi_m(q)$, and demonstrates the improvement that may be obtained compared to (8).

To guarantee a particular probability p using (8), the field size q must satisfy

$$q \geq \frac{1}{1-p^{1/m}} = \frac{1}{2} + m\log\frac{1}{p} + O\left(\frac{1}{m}\right).$$

Hence the required field size increases linearly with the size of the matrix.

Let $\pi_\infty(q) = \lim_{m\to\infty}\pi_m(q)$ then

$$\pi_\infty(q) = \prod_{i=1}^{\infty}\left(1-q^{-i}\right) = q^{1/24}\left(\frac{1}{2}\vartheta_1'\left(q^{-1/2}\right)\right)^{1/3},$$

where ϑ_1 is the Jacobi theta function [8, Equation 8.181.3] and

$$\begin{aligned}\vartheta_1'(q) &= \frac{\partial}{\partial z}\vartheta_1(z,q)\Big|_{z=0} \\ &= 2\sum_{i=0}^{\infty}(-1)^i(1+2i)q^{-\frac{1}{2}(i+\frac{1}{2})^2}.\end{aligned}$$

Truncating the latter series gives the following lower bound,

$$\pi_\infty(q) \geq \left(1-\frac{3}{x}\right)^{1/3}.$$

This lower bound is compared to π_∞ for the first 20 primes in Figure 4. For a given probability p in (7), the required field size q for $m\to\infty$ satisfies

$$q \geq \frac{3}{1-p^3}.$$

which does not depend on m.

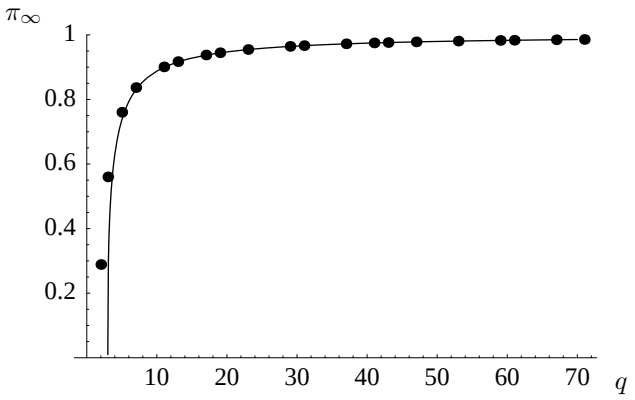

Fig. 4. Lower bound (solid line) and $\pi_\infty(q)$ (dots).

VI. Concluding remarks

Random network coding is a promising decentralized approach for multicast. One of the main implementation considerations is the size of the finite field required to achieve a specified probability that every sink can decode every source. This paper presented a new bound on the success probability, which in certain circumstances is tighter that the previous bound. We also presented a heuristic argument that motivates the investigation of tighter bounds for large random networks, based on the distribution of rank of large random finite field matrices.

Acknowledgments

This work was performed while A. Tauste Campo was visiting the Institute for Telecommunications Research. This work was supported by the Australian Government under grant DP0557310, and by the Defence Science and Technology Organisation under contracts 4500485167 and 4500550654. The authors would like to thank Ian Grivell and Terence Chan and for helpful discussions.

References

[1] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inform. Theory*, vol. 46, no. 4, pp. 1204–1216, July 2000.

[2] S.-Y. R. Li, R. Yeung, and N. Cai, "Linear network coding," *IEEE Trans. Inform. Theory*, vol. 49, no. 2, pp. 371–381, Feb. 2003.

[3] R. Koetter and M. Médard, "An algebraic approach to network coding," *IEEE/ACM Trans. Networking*, vol. 11, no. 5, pp. 782–795, Oct. 2003.

[4] T. Ho, M. Médard, J. Shi, M. Effros, and D. R. Karger, "On randomized network coding," in *41st Annual Allerton Conference on Communication, Control and Computing*, Monticello, USA, 2003.

[5] T. Ho, R. Koetter, M. Médard, D. R. Karger, and M. Effros, "The benefits of coding over routing in a randomized setting," in *IEEE Int. Symp. Inform. Theory*, Yokohama, Japan, 2003.

[6] T. Ho, M. Médard, R. Koetter, D. R. Karger, and M. Effros, "Toward a random operation of networks," *IEEE Trans. Inform. Theory*, 2004, submitted.

[7] C. Cooper, "On the distribution of rank of a random matrix over a finite field," *Random Struct. Algorithms*, vol. 17, pp. 197–212, 2000.

[8] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, 5th ed. London: Academic Press, 1994.